Proc. Natl. Acad. Sci. USA
Vol. 73, No. 7, pp. 2289–2293, July 1976
Biochemistry

DNA sequence organization in the lepidopteran *Antheraea pernyi*

(reassociation kinetics/hydroxyapatite/insect genome/interspersion)

A. EFSTRATIADIS*, W. R. CRAIN†, R. J. BRITTEN†‡, E. H. DAVIDSON†, AND F. C. KAFATOS*

* The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138; † Division of Biology, California Institute of Technology, Pasadena, Calif. 91125

Contributed by Roy Britten, April 23, 1976

ABSTRACT **A large fraction of the genome of the lepidopteran *Antheraea pernyi* consists of interspersed single-copy sequences and repetitive DNA sequences of 300 nucleotide average length. Most of the single-copy sequences are about 800 nucleotides long and a minority are considerably longer. Thus, the organization of the DNA sequences of this insect is similar to that of most higher organisms and different from that of *Drosophila*.**

Recent studies have shown that repetitive and nonrepetitive sequences are extensively interspersed in the genomes of several eukaryotes. Studies of this kind have been reported for calf (1), rat (2), *Xenopus* (3), sea urchin (4), *Aplysia*, a mollusc (5), a slime mold (6), and six other invertebrate species from diverse phylogenetic groups (7, 8). It was shown that in these organisms the major fraction of the DNA is organized in a pattern described first in detail for *Xenopus* DNA ("*Xenopus* pattern"). This pattern of DNA sequence organization has the following characteristics: (*a*) the majority of the repetitive sequences are short (with an average length of 300 nucleotides), and these alternate with single-copy sequences; (*b*) the majority of the single-copy sequences extend for less than about 1500 nucleotides before terminating in repetitive sequence elements; (*c*) a fraction of the single-copy sequences are longer, up to several thousand nucleotides in length, but are also interspersed with repetitive sequences.

A strikingly different pattern of DNA sequence organization has been found in two insects, *Drosophila melanogaster*, a dipteran (9, 10) and in the honeybee, a hymenopteran (Crain, Davidson, and Britten, in preparation). The *Drosophila* genome contains few short repetitive sequences interspersed with single copy regions (9, 10). The electron micrographic studies of Manning, Schmid and Davidson (9) indicate that the repetitive sequence regions range from a few hundred to more than 10,000 nucleotides in length, Manning *et al.* were unable to measure the length of the single-copy sequences, but they estimated that there were about 2800 repetitive regions which would have an average spacing of 30,000 nucleotides of single-copy DNA if they were generally distributed throughout the DNA. Recent measurements (Finnegan, Rubin, and Hogness, personal communication) with cloned *Drosophila* DNA sequences show that in some cases a small fraction of a repetitive region has a sequence in common with a small fraction of another region. This suggests that there is another level of sequence organization of the repeated sequences still to be revealed.

The widespread occurrence of the "*Xenopus* pattern" of DNA organization (7, 8) suggests that the short interspersed repetitive sequences might play an important role in chromosome function. It was postulated that they are involved in the regulation of genetic activity (11). This notion was indirectly supported by the observation that the sea urchin DNA single-copy sequences contiguous to interspersed repetitive elements include most or all of the structural genes that are active in embryogenesis (12). It is important, therefore, to determine whether or not the *Drosophila* and honeybee exceptions are characteristic of DNA sequence arrangement in insects. As we report here, silkmoth DNA sequence organization belongs to the more general "*Xenopus* pattern". In addition, knowledge of the general features of genome organization in the silkmoth provides a necessary background for ongoing studies of chorion genes and their organization in this species.

MATERIALS AND METHODS

Preparation of DNA. DNA was extracted from nondeveloping pupae of *Antheraea pernyi*, obtained commercially and stored at 2°. For each preparation 60 pupae were shredded for 1 min (Waring Blendor full speed, in 0.1 M NaCl, 0.01 M EDTA, 0.05 M Tris at pH 8.7). The homogenate was filtered (Nitex, 153 and 56 μm mesh) and spun for 10 min at 4000 × *g*. The pellet was resuspended in the same buffer containing 1 M $NaClO_4$ and made 2% in sodium dodecyl sulfate ($NaDodSO_4$). After 1 hr at 37° an equal volume of Sevag solution (chloroform:isoamyl alcohol, 24:1, vol/vol) was added. The suspension was shaken for 15 min and centrifuged. An equal volume of phenol was added to the aqueous phase and, after shaking, a volume of Sevag solution was added. The aqueous phase was reextracted with phenol–Sevag solution (1:1) twice. The DNA was precipitated with 0.5 volume of cold isopropanol, wound, dissolved overnight in 1 mM EDTA at pH 7, and digested for 1 hr at 37° with 50 μg/ml of RNase A (preheated for 10 min at 80°). The solution was digested with 200 μg/ml of Pronase for 1 hr at 37° (preincubated for 2 hr at 37°) and extracted with phenol–Sevag solution. The DNA was precipitated with 2 volumes of ethanol and dissolved in 1 mM EDTA at pH 7. To remove contaminating polysaccharides, the solution was spun for 30 min at 150,000 × *g*.

[^{3}H]DNA was synthesized by ovarian follicles (stages from germarium to degeneration of the nurse cells), in organ culture for 8 hr in 1.5 ml of Grace's medium containing 1 mCi of [*methyl*-^{3}H]thymidine (specific activity 50.8 Ci/mmol). The DNA was extracted from nuclei and showed a specific activity of about 2×10^4 cpm/μg.

The hyperchromicity of all DNA preparations was about 27% of the optical density after full denaturation at 98°. Melting profiles and analytical equilibrium density ultracentrifugation in CsCl gradients did not distinguish such DNA preparations from the DNA of pupal testes, which are rich in spermatocytes. Relative to a phage SP01 DNA marker the DNA had a buoyant density of 1.694 g/ml, corresponding (13) to 35% G+C.

Abbreviations: HAP, hydroxyapatite; PB, phosphate buffer; $NaDodSO_4$, sodium dodecyl sulfate; t_m, temperature at which 50% of the total change in a melting curve has occurred; C_0t, DNA concentration (moles of nucleotide/liter) × time (second).

‡ Also Staff Member, Carnegie Institution of Washington, 1530 P Street, N.W., Washington, D.C. 20005.

Fragmentation of DNA. DNA fragments of desired lengths were generated by shearing in a Virtis 60 K homogenizer (14) or were selected from sheared DNA fractionated in preparative alkaline sucrose gradients. Single-stranded DNA lengths were determined by sedimentation in isokinetic alkaline sucrose gradients (15) as described previously (5, 7).

RESULTS

Reassociation Kinetics of 250 Nucleotide DNA Fragments. The reassociation kinetics of 250 nucleotide long DNA fragments are shown in Fig. 1. Because of the wide range of kinetic components it was necessary to determine independently the rate of single-copy DNA reassociation. The open triangles show measurements of the reassociation of single-copy tracer with an excess of total 250 nucleotide long DNA driver. The least squares solution for a single second order component for these data gives a rate constant of 1.05×10^{-3} M^{-1} sec^{-1}. This is the rate constant expected for the genome size of 1 pg, which was estimated in preliminary experiments by measuring photometrically the amount of DNA in Feulgen stained nuclei of testis squashes, with *Drosophila* testis squashes as standard, and also by measuring the DNA content of counted sperm samples by diphenylamine. The circles in Fig. 1a show the reassociation of 250 nucleotide long fragments of total *Antheraea pernyi* DNA and the solid line drawn through them is the least squares solution for three second order components assuming only that the rate of the single-copy component is 1.05×10^{-3} M^{-1} sec^{-1}. The reassociation of these three individual components is portrayed by the dashed lines in Fig. 1a and the rate constants and fraction of DNA fragments given by the least squares solution are listed in Table 1. There is no deviation which suggests the presence of other components. Furthermore it is clear that since the principal part of the reassociation is initiated at about C_0t 0.05 there is a major fast component which is half reassociated at about C_0t 0.5. However there is no way to decide from these measurements whether several other components exist or even a continuum of components with repetition frequencies ranging from a few thousand copies to a few copies. We use in our analysis the minimum number of repetitive components that the reassociation kinetic measurements require, and term these the very fast or foldback fraction, and the fast and slow repetitive components; the latter two components are present in about 1600 and 15 copies, respectively.

Determination of the quantities of repetitive sequences from the magnitude of the kinetic components measured by hydroxyapatite (HAP) binding is of course affected by their interspersion. Our estimate of the fraction of the genome that is actually present in the repetitive and single-copy sequences is also shown in Table 1. The measured components may be described as: fragments which contain only single-copy sequences; fragments which contain slow sequences but not more rapidly reassociating sequences; fragments containing fast sequences but not very fast or foldback; and finally those containing very fast or foldback sequences. Each fraction is bound to HAP as a result of the reassociation of the most rapidly reassociating sequence present and may include significant quantities of those components which reassociate more slowly.

Reassociation Kinetics of 2200 Nucleotide DNA Fragments. ^{3}H-Labeled 2200 nucleotide fragments were incubated to various C_0t values with an excess of unlabeled 250 nucleotide driver DNA (tracer:driver, 1:360), and the fraction of the labeled DNA bound to HAP was measured. Fig. 1b shows the results, together with the least squares solution for two unrestricted second order components. Table 1 lists the characteristics of the two kinetic components resolved by the computer.

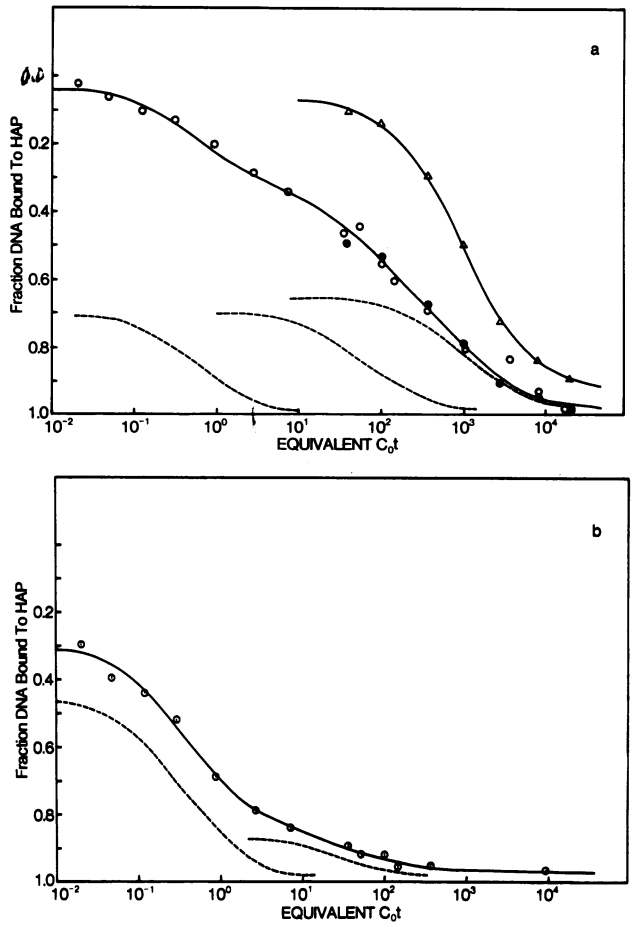

FIG. 1. Reassociation kinetics of *A. pernyi* DNA at fragment lengths of 250 nucleotides and 2200 nucleotides. DNA was reassociated at 0.35 mg/ml in 0.12 M PB at 60° ($C_0t < 0.5$) or at 7 mg/ml in 0.4 M PB at 66° ($C_0t > 0.5$). Data for reassociation in 0.4 M PB are corrected to the equivalent C_0t in 0.12 M PB. The fraction of the DNA fragments containing duplex regions was estimated by HAP binding. Fractions were eluted with 0.12 M PB at 60° (single-stranded, unbound) and 98° (duplex-containing, bound). Fig. 1a shows the reassociation of 250 nucleotide fragments of unlabeled total moth DNA (○—○). The reassociation was followed by monitoring A_{260} in the HAP column eluates. The rate of reassociation of isolated single-copy sequences is also shown (△—△). For this analysis, a fraction of 250 nucleotide tracer not bound to HAP at C_0t 3000 was reassociated together with a 75-fold mass excess of total unlabeled 250 nucleotide driver DNA (●—●). Fig. 1b shows the reassociation of ^{3}H-labeled 2200 nucleotide tracer DNA in the presence of a 360-fold excess of 250 nucleotide unlabeled DNA. The solid curves in panels a and b represent the solutions of a least squares computer analysis of the data that yielded the second order components indicated by the broken lines.

It is immediately clear that the long DNA fragments reassociate very rapidly with the short driver fragments. For example, the fraction of the 2200 nucleotide fragments which have not reassociated with short fragments by C_0t 10 is only 1/7 of the corresponding value for 250 nucleotide fragments themselves. The rapid reaction cannot be accounted for by the effect of length on the reassociation rate (16) and is indicative of DNA sequence interspersion (3, 4).

The apparent reassociation constants for the slow and fast repetitive components of our solution for the 2200 nucleotide fragments are in reasonable agreement with the corresponding values at 250 nucleotides. The rate constants for both components are somewhat greater for the 2200 nucleotide fragments.

Table 1. Reassociation analysis of *A. pernyi* DNA at two fragment lengths

Fragment length (nucleotides)	Fragments bearing exclusively single-copy sequences ("single-copy")		Fragments bearing repetitive sequences of only the slowly reassociating type ("slow")		Fragments bearing repetitive sequences of the rapidly reassociating type ("fast")		Fragments bearing foldbacks and/or very rapidly reassociating repetitive sequences ("very fast")
	Fraction of total fragments	Rate constant (M^{-1} sec^{-1})	Fraction of total fragments	Rate constant (M^{-1} sec^{-1})	Fraction of total fragments	Rate constant (M^{-1} sec^{-1})	Fraction of total fragments
250	0.35	1.05×10^{-3}	0.30	1.53×10^{-2}	0.30	1.69	0.05
2200	<0.03		0.12	2.99×10^{-2}	0.54	2.63	0.31
Estimate of fraction of genome in sequence class*	0.56		0.30		0.10		0.04

* These estimates are approximate and have been made in the following way: the single-copy fraction is the remainder after all of the repetitive classes have been calculated. For the slow repetitive sequences we use the fraction of fragments observed in the kinetics of reassociation of 250 nucleotide fragments, since we do not know their sequence organization. For estimating the quantity of the very fast fraction, we have reduced slightly the quantity observed with 250 nucleotide fragments by analogy to the results for the DNA of other species. The estimate for the fast fraction was obtained from the sum of fast plus very fast fractions as calculated in the *text*, after subtraction of the value listed in this table for the very fast fraction.

This effect is not significant for the slow component since there is too little present in the 2200 nucleotide case for an accurate least squares solution. For the fast component the apparent 50% acceleration could result from inaccuracy or because the full length of the repetitive sequence element is likely to be present on the 2200 nucleotide fragments whereas on the short fragments the sequences are most often broken as the fragments are sheared to 250 nucleotide length.

In any case the major difference between the 2200 nucleotide and 250 nucleotide solutions is in the proportions of fragments which appear to carry each class of sequence. Fragments carrying exclusively single-copy DNA account for ⅛ of the 250 nucleotide DNA (see Fig. 1a) but are undetectable at a length of 2200 nucleotides. Clearly, single-copy sequences are extensively interspersed with repetitive sequences in the *A. pernyi* genome. The large quantity of 2200 nucleotide fragments that carry fast repetitive sequences indicates that these sequences are widely dispersed in the genome. The decrease in the proportion of fragments which appear to be carrying slow repetitive sequences is clearly due to the interspersion of this class with the fast repetitive sequences.

Therefore we know that both single-copy and slow repetitive sequences are interspersed with fast repetitive sequences. However we do not know to what extent slow and single-copy sequences are interspersed with each other. It is also unresolved whether the 12% slow kinetic component in the 2200 nucleotide curve contains some linked single-copy sequences. This point is emphasized here because it relates to the following measurements of the spacing between repetitive sequences.

Measurement of Length and Amount of Repetitive DNA Sequences Using S1 Nuclease. To study the characteristics of repetitive DNA sequences, we incubated *A. pernyi* DNA fragments of about 1800 nucleotides to C_0t 10 and digested them with S1 nuclease under standard conditions (17, and Britten, Graham, Eden, Painchaud, & Davidson, submitted for publication) which leave duplex regions intact even if they contain as much as 20% mismatch. The duplexes were then separated from the partially digested single-stranded fragments by binding to HAP. The HAP bound fraction was chromatographed on agarose A-50 (Fig. 2). Approximately 75% (18% of the starting DNA) eluted almost coincidentally with a 300 nucleotide marker. The single strand length was shown by subsequent alkaline sucrose gradient centrifugation to be about 230 nucleotides. The remainder of the S1 nuclease resistant material (6% of the starting DNA) was excluded from the agarose column; it was polydispersed in alkaline sucrose gradients, and varied from a few hundred to a few thousand nucleotides in single strand length.

Fig. 3 and Table 2 show spectrophotometric melting measurements on fractions obtained in the chromatograph of Fig. 2. The t_m of the long duplex fraction, 33, is actually slightly higher than that of native DNA indicating the presence of long repetitive sequences relatively high in GC content. The t_m of the short duplexes in about 74°. About 3° reduction in t_m is expected to result from the short length of these fragments. Thus, we estimate a reduction in t_m of about 7° due to mis-

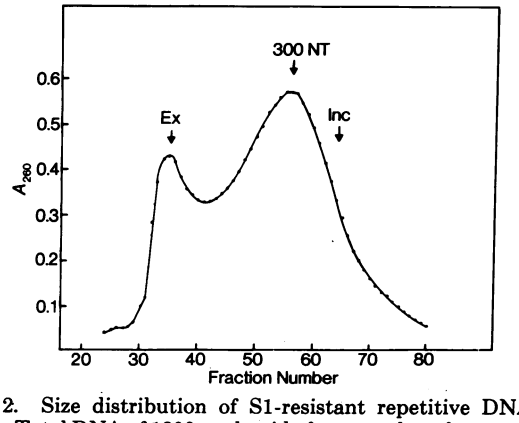

FIG. 2. Size distribution of S1-resistant repetitive DNA sequences. Total DNA of 1800 nucleotide fragment length was reassociated to C_0t 10 and treated with S1 nuclease. The enzyme-resistant duplexes were passed over HAP and eluted with 0.4 M PB. They were then chromatographed on a calibrated agarose A-50 column in 0.12 M PB. The arrows show the elution positions of an exclusion marker (Ex, long native DNA); an inclusion marker (Inc, $^{32}PO_4^{-3}$); and a 300 nucleotide long DNA marker (300 NT).

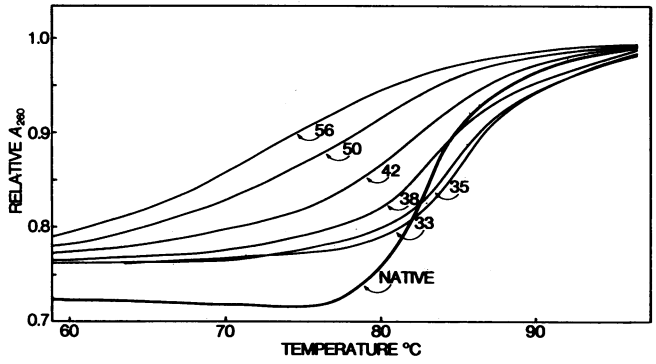

FIG. 3. Thermal stability and hyperchromicity of repetitive sequences as a function of their elution position from the A-50 column. Individual fractions (identified by number) from the A-50 column shown in Fig. 2 were melted in a water-jacketed cell in a spectrophotometer, and the A_{260} was measured. The solvent was the 0.12 M PB column eluant.

match. It follows that the short interspersed repeated sequences show greater evolutionary divergence than do the long repeated sequences. A similar result has been observed for all of the other species that have been examined in this way, including sea urchin (Britten *et al.*, submitted for publication), *Xenopus* (11, 17), calf (Britten, unpublished data), *Spisula* (7), and sea hare (Angerer and Britten, unpublished data).

The fraction of the DNA which is resistant to S1 nuclease under these conditions (24%) permits an estimate of the actual fraction of the DNA which is in repetitive sequences. Approximately 10% of the S1-resistant material might be single strand tails, as suggested by the reduced hyperchromicity (see Table 2). Therefore, the repetitive sequences reassociated at C_0t 10 are approximately 22% of the total DNA. At C_0t 10 the fast fraction would have been completely reassociated while the slow fraction would only be about 25% reassociated (Table 1). Therefore, the maximum amount of slow repeated DNA in the nuclease resistant fraction is one-fourth of 30% or about 8% of the genome. The fast and very fast fractions are completely reassociated at C_0t 10 and correspond to about 14% of the genome.

Table 2. Thermal stability of S1-resistant repetitive sequences

Fraction of the A-50 column (Fig. 2)	Hyperchromicity*	t_m (°C)
33	0.23	85.7
35	0.23	85.0
38	0.23	84.0
42	0.22	81.2
50	0.21	77.2
56	0.20	74.2
Native DNA	0.27	84.0

* Hyperchromicity is calculated as (A_{260} at 98° − A_{260} at 60°)/A_{260} at 98°. As can be seen from Fig. 2 the agarose column contributes an unknown small UV absorbing background and these figures are probably too low by about 5 or 10% of their values. The average hyperchromicity of the S1-resistant material compared to that of fully paired DNA, suggests the existence of 15–20% single-stranded sequence. Approximately half of this is due to mismatch, according to the t_m measurements, suggesting that the other half is single strand tails.

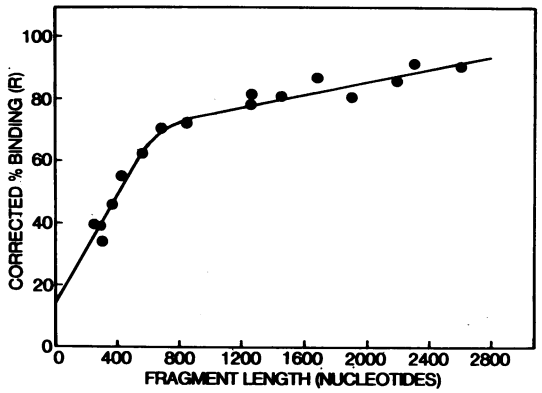

FIG. 4. The fraction of DNA fragments containing repetitive sequence elements as a function of fragment size. Tracer fragments of the indicated length were reassociated together with a 400-fold excess of 250 nucleotide long driver, at driver C_0t 20. *R* represents the fraction of tracer bound, after correction for the binding that occurred at C_0t 5×10^{-4}. The line is an unrestricted computer least squares fit. For further details, see the *text*.

The Spacing of Repeated Sequences as Estimated from the Increase with Length of the Fraction of Fragments Containing Repeated Sequences. Tracer fragments of various defined lengths were obtained from sheared DNA fractionated on a preparative alkaline sucrose gradient. Each fraction was reassociated with a 400-fold excess of 250 nucleotide unlabeled driver DNA, at driver C_0t 20, and the fraction of tracer fragments containing reassociated sequences was assayed by HAP chromatography. At C_0t 20 the fast fraction is completely reassociated and the single-copy fraction is not reassociated. The slow fraction is partially reassociated. Thus, the HAP binding of the tracer fragments is due principally to the reassociation of the fast fraction with some contribution from the slow sequences and none from the single-copy sequences. As a result of interspersion of different sequence classes with each other the length of the single stranded regions linked to the reassociated sequences increases with fragment length.

Fig. 4 shows our best estimate of *R*, the fraction of fragments which contain repeated sequences as a function of tracer fragment length (3, 4). For this estimate a correction for the increase in binding at C_0t 5×10^{-4} due to foldback and very fast repeated sequences (3, 4) was made using the formula $R = F - Z/1 - Z$, where *F* is the fraction of tracer fragment bound to HAP at C_0t 20 and Z is the fraction of fragments bound at C_0t 5×10^{-4}.

The solid line of Fig. 4 represents the least squares solution of these measurements by the method described by Graham *et al.* (4). This curve extrapolates to 14% for very short fragment lengths. This is surely too low an estimate of the total amount of repetitive sequences reassociated by C_0t 20. The solution is obviously affected by scatter in the data at short fragment lengths. Many other solutions are possible and the differences between them are unimportant in their effect on the interpretation of the major features of the curve in Fig. 4. All solutions show a strong decrease in slope at about 800 nucleotides and a relatively small slope at long fragment lengths. The interpretation of such "*R*" curves previously described (3, 4) indicates that there are many repetitive sequence elements which are spaced by about 800 nucleotides of DNA which was not reassociated at C_0t 20. In addition there is another class of repetitive sequences which are spaced much farther apart. Since no curvature is evident in the data at long lengths we conclude that the longer spacings are mostly greater than 2400 nucleotides.

Thus, we conclude that there are two broad classes of repetitive sequence spacing in *A. pernyi* DNA which can be termed long period and short period interspersion. We can also conclude that almost all of the single-copy DNA sequences are included in this interspersion pattern. We presume by analogy to the situation in the DNA of other species, that many of these single-copy sequences have a length of about 800 nucleotides and that another class is longer. However this pattern is not directly demonstrated for *A. pernyi* DNA due to the presence of substantial amounts of slow repetitive component, which is only partially reassociated at C_0t 20 in these measurements. The slow repetitive sequences could also be interspersed with fast repetitive sequences and thus the 800 nucleotide lengths may include slow in addition to single-copy sequences. For this reason it is impossible to derive an accurate conclusion about the preponderant length of the interspersed single-copy sequences. The principal part of the DNA which is reassociated at C_0t 20 is the fast repetitive fraction, and most of the 800 nucleotide long spaces separate fast repetitive sequences. The main conclusions of this work are nevertheless that most single-copy sequences in *A. pernyi* DNA are interspersed with repetitive DNA sequences, and that the repetitive DNA sequences are typically 200 to 300 nucleotides in length.

DISCUSSION

The fraction of DNA in the several kinetic components, summarized in Table 1, does not reflect accurately their abundance in the genome, due to sequence interspersion. Our best estimate of the amount of DNA in the various sequence classes given in Table 1, are derived from measurements of S1 nuclease resistance. The kinetics of reassociation of labeled fragments that are 2200 nucleotide long with short driver DNA indicate that almost all of the single-copy DNA sequence and most of the slow repeated DNA sequence are interspersed with fast repetitive sequences. In fact, greater than 85% of 2200 nucleotide fragments contain repetitive sequences which have reassociated by C_0t 10.

The length and melting characteristics of S1 nuclease-resistant duplexes show a striking similarity to those for many other species. The short repetitive sequence duplexes account for about ¾ of the total enzyme-resistant repetitive DNA at C_0t 10, and have a mode length between 200 and 300 nucleotides. About 25% of the S1-resistant duplexes (6% of the genome) are excluded from agarose A-50 and many of these are reassociation products from longer repetitive regions in the genome. They appear to be polydispersed in alkaline sucrose sedimentation measurements and little can yet be said about their actual sequence organization. The measurements shown in Fig. 4 of the fraction of fragments which contain repeated sequences as a function of length show that the sequence organization of *A. pernyi* is similar to that of *Xenopus laevis*. A very large fraction of the single-copy DNA is interspersed with repetitive sequences. The possibility exists that the slow, fast, and single-copy sequences are all interspersed with each other and this prevents a direct conclusion from the "*R*" curve measurements of the length of the single-copy sequences. However it is clear that most of the fast repetitive sequences are short (200–300 nucleotides) and spaced by about 800 nucleotides, partially if not exclusively with single-copy DNA sequences.

In conclusion, we have shown that the *A. pernyi* genome is organized in a pattern quantitatively similar to the patterns of DNA sequence organization in *Xenopus laevis*, *Strongylocentrotus purpuratus*, and *Aplysia californica*. Therefore, the widespread "*Xenopus* pattern" occurs even among insects. After this study was completed we showed that the "*Xenopus* pattern" also occurs in another insect, *Musca domestica*, a dipteran (Crain *et al.*, in preparation). The "Drosophila pattern" thus seems exceptional, as it has been observed only in two insects, the honeybee and *Drosophila* itself (9, 10). The functional and evolutionary implications of this unusual sequence organization are yet unclear. The pattern of sequence organization revealed by these studies in the genome of *A. pernyi*, the first lepidopteran to be studied, is of very general occurrence throughout the animal kingdom as well as in some plants. We attribute the general occurrence and evolutionary conservatism of this pattern to some, as yet unknown, crucial function in the genome.

We thank Drs. F. Eden and R. Angerer for helpful discussions, M. Smart for analytical CsCl ultracentrifugation, Dr. J. Nardi for photometric estimates of genome size, Dr. J. Sheppard for counted sperm samples, L. DeLong for valuable technical assistance, and M. Koehler for help with the figures. This work was supported by Grants NSF (BMS72-02336) and NIH (2ROI-HD 04701-06) to F.C.K., and by NIH Grants (HD-05753) and (GM-20927) and by NSF Grant (BMS75-07359) to R.J.B. and E.H.D. W.R.C. holds a postdoctoral fellowship from the National Institutes of Health (GM-55726).

1. Britten, R. J. & Smith, J. F. (1970) *Carnegie Inst. Washington Yearbk.* **68**, 378–386.
2. Bonner, J., Garrard, W. T., Gottesfeld, J., Holmes, D. S., Sevall, J. S. & Wilkes, M. M. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 303–310.
3. Davidson, E. H., Hough, B. R., Amenson, C. S. & Britten, R. J. (1973) *J. Mol. Biol.* **77**, 1–23.
4. Graham, D. E., Neufeld, B. R., Davidson, E. H. & Britten, R. J. (1974) *Cell* **1**, 127–137.
5. Angerer, R. C., Davidson, E. H. & Britten, R. J. (1975) *Cell* **6**, 29–39.
6. Firtel, R. A. & Kindle, K. (1975) *Cell* **5**, 401–413.
7. Goldberg, R. B., Crain, W. R., Ruderman, J. V., Moore, G. P., Barnett, T. R., Higgins, R. C., Gelfand, R. A., Galau, G. A., Britten, R. J. & Davidson, E. H. (1975) *Chromosoma* **51**, 225–251.
8. Davidson, E. H., Galau, G. A., Angerer, R. C. & Britten, R. J. (1975) *Chromosoma* **51**, 253–259.
9. Manning, J. E., Schmid, C. W. & Davidson, N. (1975) *Cell* **4**, 141–155.
10. Crain, W. R., Eden, F. C., Pearson, W. R., Davidson, E. H. & Britten, R. J. (1976), *Chromosoma*, in press.
11. Davidson, E. H. & Britten, R. J. (1973) *Q. Rev. Biol.* **48**, 565–613.
12. Davidson, E. H., Hough, B. R., Klein, W. H. & Britten, R. J. (1975) *Cell* **4**, 217–238.
13. Mandel, M., Schildkraut, C. L. & Marmur, J. (1968) in *Methods in Enzymology*, eds. Grossman, L & Moldave, K. (Academic Press, New York), Vol. 12B, pp. 184–195.
14. Britten, R. J., Graham, D. E. & Neufeld, B. R. (1974) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 29, pp. 363–418.
15. Noll, H. (1967) *Nature* **215**, 360–363.
16. Wetmur, J. G. & Davidson, N. (1968) *J. Mol. Biol.* **31**, 349–370.
17. Davidson, E. H., Graham, D. E., Neufeld, B. R., Chamberlin, M. E., Amenson, C. S., Hough, B. R. & Britten, R. J. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 295–301.